

Requirements for a successful defence reaction by the CRISPR–Cas subtype I-B system

Britta Stoll*, Lisa-Katharina Maier*, Sita J. Lange†, Jutta Brendel*, Susan Fischer*, Rolf Backofen†‡§|| and Anita Marchfelder*¹

*Biology II, Ulm University, Albert-Einstein-Allee 11, 89069 Ulm, Germany, †Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany, ‡Centre for Biological Systems Analysis (ZBSA), University of Freiburg, Habsburgerstrasse 49, 79104 Freiburg, Germany, §Centre for Biological Signalling Studies (BIOS), Cluster of Excellence, University of Freiburg, Schänzlestrasse 18, 79104 Freiburg, Germany, and ||Centre for Non-Coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, 1870 Frederiksberg C, Denmark

Abstract

Uptake of foreign mobile genetic elements is often detrimental and can result in cell death. For protection against invasion, prokaryotes have developed several defence mechanisms, which take effect at all stages of infection; an example is the recently discovered CRISPR (clustered regularly interspaced short palindromic repeats)–Cas (CRISPR-associated) immune system. This defence system directly degrades invading genetic material and is present in almost all archaea and many bacteria. Current data indicate a large variety of mechanistic molecular approaches. Although almost all archaea carry this defence weapon, only a few archaeal systems have been fully characterized. In the present paper, we summarize the prerequisites for the detection and degradation of invaders in the halophilic archaeon *Haloferax volcanii*. *H. volcanii* encodes a subtype I-B CRISPR–Cas system and the defence can be triggered by a plasmid-based invader. Six different target-interference motifs are recognized by the *Haloferax* defence and a 9-nt non-contiguous seed sequence is essential. The repeat sequence has the potential to fold into a minimal stem-loop structure, which is conserved in haloarchaea and might be recognized by the Cas6 endoribonuclease during the processing of CRISPR loci into mature crRNA (CRISPR RNA). Individual crRNA species were present in very different concentrations according to an RNA-Seq analysis and many were unable to trigger a successful defence reaction. Recognition of the plasmid invader does not depend on its copy number, but instead results indicate a dependency on the type of origin present on the plasmid.

The diversity of CRISPR (clustered regularly interspaced short palindromic repeats)–Cas (CRISPR-associated) systems

A recently discovered defence strategy of prokaryotes against foreign genetic elements is the CRISPR–Cas system [1–6]. Central elements of this system are the Cas proteins and the crRNAs (CRISPR RNAs). The system detects invading foreign nucleic acid and degrades it. The CRISPR–Cas system has been identified in almost all archaea and in approximately half of bacteria. Comparison of the Cas proteins from different organisms revealed that they can be sorted into 45 Cas protein families [7] that have been grouped into three major types (I–III), which show considerable differences between each other [8]. The major types have been further divided into ten subtypes (I–A–I–F, II–A and II–B, and III–A and III–B), which again show significant variances. The major Type I contains the most subtypes and these six subtypes are not equally distributed between the Bacteria

and Archaea domains [9]. Archaea contain mainly subtypes I–A, I–B and I–D, whereas bacteria contain mainly subtypes I–C, I–E and I–F [9]. Comparably little is known about the archaeal CRISPR–Cas systems, especially about the subtypes I–B and I–D [10]. In the present paper, we summarize our investigations concerning the subtype I–B in the halophilic archaeon *Haloferax volcanii*.

Characteristics of the *Haloferax* CRISPR–Cas subtype I–B system

H. volcanii is a model archaeon, representative of the euryarchaeal class of haloarchaea [11]. It is halophilic, requiring high salt concentrations for growth and contains similar concentrations of salt intracellularly to cope with the high salt concentration in the medium. *H. volcanii* (H119) encodes a subtype I–B system with eight Cas proteins (Cas1–Cas8b) and three CRISPR RNAs [12,13]. We have shown previously that the CRISPR–Cas system is still active, processing all three CRISPR loci, which are constitutively expressed and processed into mature crRNAs [14]. The spacers encoded in the three CRISPR loci show only two matches to sequences in the public sequence databases: one with an overall sequence identity of 76 % with the *Haloferax* genome itself and a second one with a sequence identity

Key words: clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas), CRISPR RNA (crRNA), *Haloferax volcanii*, protospacer-adjacent motif (PAM), subtype I–B CRISPR, target-interference motif (TIM).

Abbreviations used: Cas, CRISPR-associated; CRISPR, clustered regularly interspaced short palindromic repeats; crRNA, CRISPR RNA; PAM, protospacer-adjacent motif; SAM, spacer-acquisition motif; TIM, target-interference motif.

¹To whom correspondence should be addressed (email anita.marchfelder@uni-ulm.de).

of 88% with an environmental sequence from a sample isolated from Lake Tyrrell, an Australian salt lake [14]. The low number of matches may be due to the fact that only a few virus sequences are present in the databases and that the *H. volcanii* strain was isolated 30 years ago and virus populations have evolved since then [14]. Analysis of the repeat structure of the three *Haloferax* CRISPR loci showed that the repeats are 30 nt in length and the sequences are identical between the three CRISPR loci except for a single nucleotide (Figure 1A). The CRISPR loci are processed into crRNAs, which are the central elements of the CRISPR–Cas defence since they direct the degradation complex to the invader in a sequence-specific manner. They contain the spacer sequence (Figure 1B) in between two repeat sequence tags that is derived from a previously encountered invader and is used to detect the invader during a repeated invasion. The different CRISPR–Cas types have developed various mechanisms to generate the functional crRNA molecule from the crRNA precursor [15]. In Type I and III systems the crRNA is processed from the crRNA precursor by representatives of the family of the Cas6 protein. Cas6 proteins have been analysed in detail from subtype I-B, I-E, I-F and III-B systems [16]. Although they catalyse the same reaction, they show very little sequence similarity and the catalytic site is made up differently [16]. The Cas6 protein in *H. volcanii* is also essential for crRNA production, since the deletion of the *cas6* gene resulted in loss of mature crRNAs (J. Brendel, B. Stoll, S.J. Lange, L.-K. Maier, R. Backofen and A. Marchfelder, unpublished work).

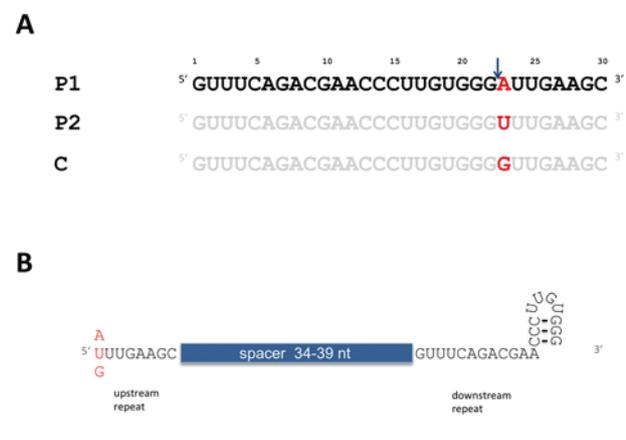
crRNA repeats and crRNA characteristics

Analysis of mature crRNAs in *Haloferax* showed that they contain an 8-nt 5' handle [17], similar to crRNAs in other organisms and systems [15] (Figure 1B). According to the crRNA sizes detected in Northern blots, *Haloferax* crRNAs have an overall length of approximately 65 nt, suggesting that they contain 22 nt of the repeat at the 3' end [14] (Figure 1B). This type of structure is similar to the crRNA structure found in subtypes I-A, I-E and I-F [15], but, interestingly, is different from the crRNA structure found in two other subtype I-B systems, those of *Methanococcus maripaludis* and *Clostridium thermocellum* [18]. *M. maripaludis* and *C. thermocellum* crRNAs are further trimmed at the 3' end as observed in subtype III-B systems [19] leaving only a few nucleotides of the downstream repeat [18]. Our observations, together with the one from the other subtype I-B systems, suggest that even within a subtype differences in the molecular details of the reaction can exist. A detailed analysis of the nature of the crRNA 3' end in *Haloferax* will show how much the crRNA characteristics in the subtype I-B systems differ.

Analysis of repeat sequences of the different CRISPR–Cas types showed that they have highly variable structures [20,21] (CRISPRmap web server: <http://rna.informatik.uni-freiburg.de/CRISPRmap>). The crRNAs from the subtypes I-E and I-F form a stable hairpin structure which

Figure 1 | The repeat sequences and crRNAs of *Haloferax*

(A) *Haloferax* CRISPR loci contain 30-nt-long repeat sequences. The repeat sequences of all three CRISPR loci are identical except for one nucleotide. The repeat sequence of the CRISPR locus P1 is shown in bold: at position 23 (where locus P1 contains an A) repeats from locus P2 contain a U, whereas repeats from locus C contain a G. The arrow depicts the position where the repeat is cleaved. (B) *Haloferax* crRNA. *Haloferax* crRNAs contain a spacer of 34–39 nt in length as the central crRNA element and an 8-nt 5' handle, which is derived from the upstream repeat. The first nucleotide of the crRNA differs between loci as shown in (A). According to Northern blot analyses, the 3' end probably consists of 22 nt (derived from the downstream repeat), but the exact nature of the 3' end still has to be determined. The repeat has the potential to form a minimal stem–loop consisting of three base pairs and a 4-nt loop. This minimal structure is conserved throughout the haloarchaea, suggesting a conserved function.



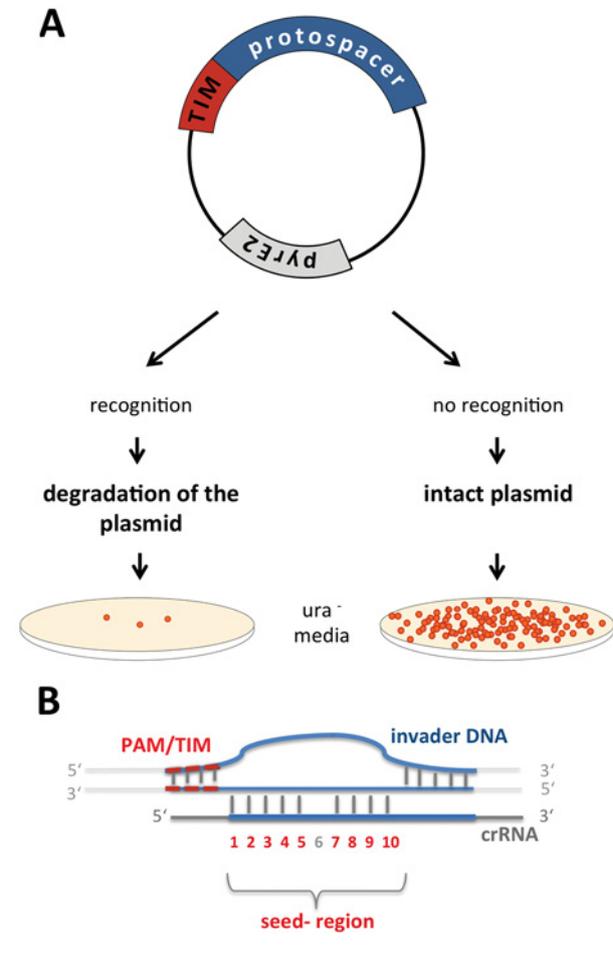
is a critical feature for recognition and cleavage by the Cas6 protein [22]. Similarity searches with the *Haloferax* repeat sequence against microbial databases showed that the repeat sequence is conserved in haloarchaea with the potential to form a minimal stem–loop structure with a 3-nt stem and a 4-nt loop (Figure 1B). This structure is conserved throughout the subtype I-B-containing haloarchaea, suggesting that it is important for function [17]. Binding by proteins might stabilize this minimal structure and facilitate cleavage at the 3' end of the stem by the Cas6b protein. Such a stabilization of a minimal stem–loop structure by the Cas6 protein has recently been shown for one of the *Sulfolobus solfataricus* Cas6 proteins and its respective repeat [22].

A plasmid invader triggers the immune system and reveals essential requirements for interference

To determine the requirements for a successful defence reaction we employed a plasmid-based invader system [14,17,23] (Figure 2A). To trigger a defence reaction, the invader must match one of the spacers in the crRNAs and contain a distinct sequence motif of approximately 2–5 nt in length, the so-called PAM (protospacer-adjacent motif) [8,24,25] (Figure 3). This motif is vital for two stages of the

Figure 2 | A plasmid invader for triggering the immune defence

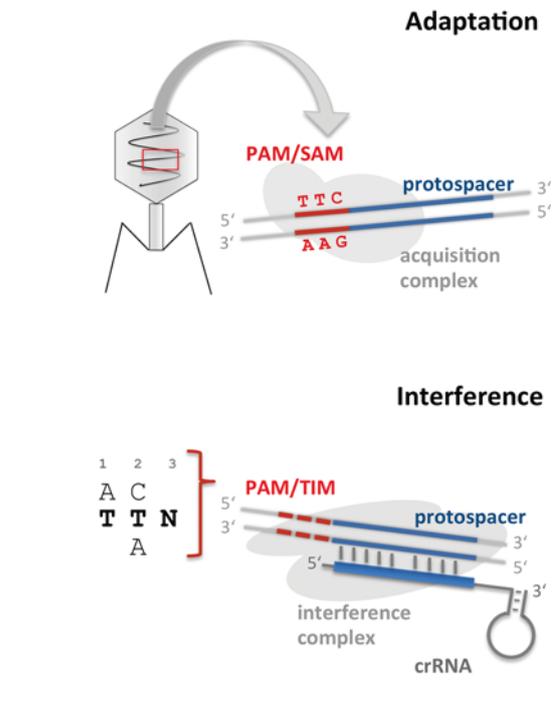
(A) Invader plasmids were constructed that carry a protospacer sequence and a trinucleotide sequence adjacent to the protospacer, i.e. the TIM. The invader plasmid contains a selection marker (*pyrE2*) to allow growth on uracil-free medium (*ura⁻* media). Only cells that contain the plasmid are able to grow on the selection medium. The defence reaction is triggered by a valid TIM sequence and recognition of the protospacer sequence by the crRNA. If the plasmid is degraded, it is indicated by a drastically reduced transformation rate (at least 100-fold). (B) A seed sequence is a prerequisite for a successful defence. The crRNA must base pair with the target over a 10-nt stretch with only one mismatch at position six allowed.



defence reaction: (i) during adaptation when a piece of the invader DNA is selected for integration into the CRISPR locus; and (ii) in interference when the invader DNA is recognized and degraded. Studies with different organisms and different CRISPR–Cas types have shown that the motifs for these two stages are not identical [25]. The motif required for selection during adaptation is more conserved than the one for the interference reaction, allowing more invader sequences to be recognized. Therefore the two motifs have been termed differently: SAM for spacer-acquisition motif and TIM for target-interference motif [25] (Figure 3).

Figure 3 | Different recognition motifs for different steps of the defence: PAM, SAM and TIM

The sequence that is selected to be integrated into the CRISPR locus for future identification of the invader has been termed the protospacer. After integration into the CRISPR locus, it is termed the spacer. Located adjacent to the protospacer sequence in the invader DNA is a signature sequence, the PAM. This short motif is 2–5-nt-long and varies between different CRISPR–Cas systems and organisms. It is important for two steps of the defence: the adaptation step and the interference step. In the adaptation step, it is important for the selection of the protospacer that is integrated into the CRISPR locus. In the interference step, it is important for recognition of the invader. Investigations into different organisms with different CRISPR–Cas systems have shown that the PAM sequence differs depending on the defence step. Therefore the motifs for the two different steps were given distinct names [25]. Requirements for the motif in the adaptation step seem to be much more stringent, thus the motif for this step was termed SAM. However, the defence reaction is a little bit more flexible: more sequences are recognized to allow a better defence. The motif for this step was called TIM.



Using the plasmid invader system, we determined the TIMs for efficient recognition by the *Haloflex* defence system [14]. In *Haloflex* the TIM sequences are 3-nt-long and located upstream of the protospacer, and we found six different TIM sequences which were effective in triggering the defence reaction: ACT, TTC, TAA, TAT, TAG and CAC.

Investigation of the prerequisites for invader detection revealed that the spacer sequence in the crRNA must form base pairs with the corresponding target sequence at a seed sequence spanning 10 nt at its 5' end that allows only a single mismatch at position six (Figure 2B). This is similar to the seed sequence detected in *Escherichia coli* and

Pseudomonas aeruginosa that was an essential prerequisite for the interference reaction [26,27]. In *E. coli*, the required sequence is only 7-nt-long with a mismatch tolerated at position six [26].

An RNA-Seq analysis of the quantities of individual crRNA species in *Haloflex* showed that they are not present in equal concentrations [17]; similar observations were previously made in other organisms [18,28–30]. This might be due to technical biases (e.g. some crRNAs are better ligated and amplified in preparation for deep sequencing), but could also be due to different stabilities of the crRNA molecules depending on the spacer sequences contained. To investigate whether all crRNAs have the same effectivity in triggering invader degradation (independent of crRNA concentration), several invaders with sequences against different spacers were tested in the plasmid invader assay. Only some of the selected crRNAs were active in triggering a defence reaction, whereas many were not. It is unclear at this point which factors influence the effectivity of a crRNA: the concentration of the crRNA, the length of the spacer, the percentage of G/C nucleotides or other additional factors that might influence the stability of the crRNA or the crRNA–target interaction.

In initial plasmid invader assays, different vectors were used, revealing that a plasmid with a high copy number was not active in triggering the defence reaction. Further analysis showed it was not the copy number, but the nature of the origin of replication that was the cause of the loss of defence. Although plasmids with an ORC (origin-recognition complex)-based mode of replication [31,32] activated the defence and were degraded, plasmids with a distinct replication mode (presumably Rep-dependent [33,34]) were not degraded [17]. Further analyses will show whether these observations are due to sterical problems (the origin is located close to the protospacer sequence on the invader plasmid) or interactions of the CRISPR–Cas system with the replication of the invader.

Conclusion

Using a plasmid-based invader we have determined that the *Haloflex* subtype I-B system recognizes six different TIMs and requires a 9-nt-long non-contiguous seed interaction between crRNA and target. A minimal stem–loop structure, which is conserved in haloarchaea, can fold in the repeat and might be important for processing by the Cas6 protein. The crRNAs in *Haloflex* are not all active in triggering the defence reaction and they appear to be present in different concentrations. Results indicate that the type of origin present on the invader plasmid may be important for a successful defence of the invader.

Acknowledgement

We thank all of the members of the DFG Research Unit FOR1680 for helpful discussions.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft in the frame of the priority programme ‘Unravelling the prokaryotic immune system’ [number FOR1680].

References

- Al-Attar, S., Westra, E.R., van der Oost, J. and Brouns, S.J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392**, 277–289
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712
- Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, 273–297
- Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71
- Garrett, R.A., Vestergaard, G. and Shah, S.A. (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.* **19**, 549–556
- Marchfelder, A., Fischer, S., Brendel, J., Stoll, B., Maier, L.K., Jäger, D., Prasse, D., Schmitz, R. and Randau, L. (2012) Small RNAs for defence and regulation in archaea. *Extremophiles* **16**, 685–696
- Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477
- Staals, R.H.J. and Brouns, S.J. (2013) Distribution and mechanism of the type I CRISPR-Cas system. In *CRISPR-Cas Systems* (Barrangou, R. and van der Oost, J., eds), pp. 145–169, Springer-Verlag, Berlin
- Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J. and van der Oost, J. (2012) The CRISPRs, they are a-changin’: how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46**, 311–339
- Hartman, A.L., Norais, C.d., Badger, J.H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T.M. et al. (2010) The complete genome sequence of *Haloflex volcanii* DS2, a model archaeon. *PLoS ONE* **5**, e9605
- Maier, L.K., Fischer, S., Stoll, B., Brendel, J., Pfeiffer, F., Dyall-Smith, M. and Marchfelder, A. (2012) The immune system of halophilic archaea. *Mob. Genet. Elements* **2**, 228–232
- Maier, L.K., Stoll, B., Brendel, J., Fischer, S., Pfeiffer, F., Dyall-Smith, M. and Marchfelder, A. (2013) The ring of confidence: a haloarchaeal CRISPR/Cas system. *Biochem. Soc. Trans.* **41**, 374–378
- Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M. and Marchfelder, A. (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biol. Chem.* **287**, 33351–33365
- Charpentier, E., van der Oost, J. and White, M.F. (2013) crRNA biogenesis. In *CRISPR-Cas Systems* (Barrangou, R. and van der Oost, J., eds), pp. 115–144, Springer-Verlag, Berlin
- Reeks, J., Sokolowski, R.D., Graham, S., Liu, H., Naismith, J.H. and White, M.F. (2013) Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochem. J.* **452**, 223–230
- Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., Backofen, R. and Marchfelder, A. (2013) Essential requirements for the detection and degradation of invaders by the *Haloflex volcanii* CRISPR/Cas system I-B. *RNA Biol.* **10**, 5
- Richter, H., Zoepfel, J., Schermuly, J., Maticzka, D., Backofen, R. and Randau, L. (2012) Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **40**, 9887–9896
- Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M. and Terns, M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956

- 20 Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61
- 21 Lange, S.J., Alkhnbashi, O.S., Rose, D., Will, S. and Backofen, R. (2013) CRISPRmap: an automated classification of repeat conservation in prokaryotic immune systems. *Nucleic Acids Res.* **41**, 8034–8044
- 22 Shao, Y. and Li, H. (2013) Recognition and cleavage of a nonstructured CRISPR RNA by its processing endonuclease Cas6. *Structure* **21**, 385–393
- 23 Gudbergdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q. and Garrett, R.A. (2011) Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49
- 24 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740
- 25 Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**, 5
- 26 Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10098–10103
- 27 Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J. and Doudna, J.A. (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10092–10097
- 28 Nickel, L., Weidenbach, K., Jager, D., Backofen, R., Lange, S.J., Heidrich, N. and Schmitz, R.A. (2013) Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol.* **10**, 5
- 29 Scholz, I., Lange, S.J., Hein, S., Hess, W.R. and Backofen, R. (2013) CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS ONE* **8**, e56470
- 30 Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V. et al. (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* **45**, 303–313
- 31 Delmas, S., Shunburne, L., Ngo, H.P. and Allers, T. (2009) Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet.* **5**, e1000552
- 32 Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet.* **3**, e77
- 33 Woods, W.G. and Dyll-Smith, M.L. (1997) Construction and analysis of a recombination-deficient (*radA*) mutant of *Haloferax volcanii*. *Mol. Microbiol.* **23**, 791–797
- 34 Charlebois, R.L., Lam, W.L., Cline, S.W. and Doolittle, W.F. (1987) Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaeobacterium. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8530–8534

Received 3 June 2013
doi:10.1042/BST20130098